PL-TR-96-2157

STATISTICAL METHODS FOR AUTOMATED DETECTION OF EXTREME EVENTS

H.L. Gray
W.A. Woodward
S.R. Sain
W.H. Frawley

Southern Methodist University
Department of Statistical Science
Dallas, TX 75275-0235

July 1996

Scientific Report No. 2

PHILLIPS LABORATORY
DIRECTORATE OF GEOPHYSICS
AIR FORCE MATERIEL COMMAND
HANSCOM AIR FORCE BASE, MA 01731-3010

SPONSORED BY
Advanced Research Projects Agency (DoD)
Nuclear Monitoring Research Office
ARPA ORDER No. 0325

MONITORED BY
Phillips Laboratory
CONTRACT No. F19628-93-C-0098

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either express or implied, of the Air Force or U.S. Government.

This technical report has been reviewed and is approved for publication.

DELAINE R. REITER
Contract Manager
Earth Sciences Division

JAMES F. LEWKOWICZ
Director
Earth Sciences Division

This report has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify PL/IM, 29 Randolph Road, Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No: 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1996	Scientific No.2

4. TITLE AND SUBTITLE

Statistical Methods for Automated Detection of Extreme Events

5. FUNDING NUMBERS

PE 62301E
PR NM95
TA GM
WU AB
F19628-95-C-0098

6. AUTHOR(S)

H.L. Gray, W.A. Woodward, S.R. Sain, W.H. Frawley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Southern Methodist University
Department of Statistical Science
Dallas, Texas 75275-0235

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Phillips Laboratory
29 Randolph Road
Hanscom AFB, MA 01731-3010
Contract Manager: Dr. Delaine Reiter/GPE

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

PL-TR-96-2157

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In several previous reports, the authors have considered the problem of outlier detection. In this report, several issues are addressed which are necessary to make our previous results broadly applicable in automated operational environments. This report basically consists of Appendix 1 and Appendix 2. The first appendix extends previous outlier detection methodology using the generalized likelihood function in such a manner that ground truth is no longer required. If the data vector contains several features, say for example 10, and several stations, say 8, observe the event, then the generalized likelihood ratio approach would require an estimate of an 80 x 80 covariance matrix, which will most likely not be feasible. In the second appendix, this problem is solved by compressing the feature vector in such a way that the detection probabilities are still optimal and estimation of a large covariance matrix is no longer required. For instance, in the example suggested in the above, the 80 x 80 matrix would be replaced by an 8 x 8 matrix. Simulation studies show no loss in this method over the previous one.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Outlier detection, data compression, likelihood ratio tests, cluster analysis	56
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

CONTENTS

Appendix A

Outlier Detection when Training Data are Unlabeled

Stephan R. Sain, H. L. Gray, and Wayne A. Woodward

Department of Statistical Science

Southern Methodist University

ABSTRACT

We consider the difficult task of using seismic signals (or any other discriminants) for detecting nuclear explosions from the large number of background signals such as earthquakes and mining blasts. Wang et al. (1996) attack the problem in terms of outlier detection, i.e. modeling the background as a mixture distribution and looking for outliers (nuclear events) from that mixture. However, those authors only considered the case in which at least some fraction of the training sample was labeled, i.e. ground truth was available, and the number of distinct classes of events was known. In the current report, we extend these results to the case in which no events in the training sample are labeled and also to the case in which the number of event types represented in the training sample is unknown. In order to accomplish this task, a preliminary clustering of the training sample data is necessary. We also briefly consider the case in which some observations in the training sample and in the potential outlier are missing.

1. Introduction

Monitoring a comprehensive test ban treaty involves the difficult problem of differentiating the seismic signal of nuclear events from the large number of seismic signals of earthquakes, mining explosions, etc. This problem is made even more difficult due to the lack of information concerning the behavior of nuclear signals in many regions of the earth. The distinguishing characteristics of small nuclear explosions are regional in nature. Therefore, the features that characterize such events are not transportable from region-to-region around the world. Certainly, in many regions there is no previous data on nuclear tests. Furthermore, in many regions, little information is available on non-nuclear events.

Wang, Woodward, Gray, Wiechecki, and Sain (1996) frame the problem of detecting nuclear events in terms of detecting outliers (nuclear events) from a mixture population (earthquakes, mining explosions, etc.) Specifically, the authors assume that the training data is a sample of size n from a mixture distribution whose density is given by

$$f(x) = \sum_{i=1}^{m} p_i g_i(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{1}$$

where m is the number of components in the mixture, $g_i(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the density associated with the ith component, the p_i, $i = 1, \ldots, m$ are the mixing proportions, and $\boldsymbol{x}$ is a d-dimensional vector of feature variables. A typical scenario might be the case in which the mixture population consists of events associated with earthquakes and mining explosions. The authors developed a modified likelihood ratio test that required no distributional assumptions concerning the outlier distribution. This is a useful practical solution because of the lack of regional training samples for nuclear events. Using the bootstrap to model the distribution of the test statistic and calculate critical values, the authors showed that this test has essentially as high a detection probability as the standard

likelihood test in which complete information concerning the distribution of the outlier population is known.

However, Wang et al. (1996) made assumptions concerning the training sample that may not be appropriate in practice. Specifically, it was assumed that the associated source component population is identifiable for $n_L \leq n$ members of the training sample where $n_L > 0$. Letting n_i denote the number of labeled (i.e. the source of the event is known) observations associated with component i, the authors assumed that the n_i, $i = 1, \ldots, m$ are random variables following a multinomial distribution and that they contain information about the mixing proportions. The parameters are estimated via the EM algorithm (see McLachlan, 1982, Redner and Walker, 1984, and Dempster, Laird, and Rubin, 1977), and this algorithm further assumes that each n_i is sufficiently large to provide initial estimates of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$.

In practice, no labeled (ground truth) data may be available, and it may well be the case that we may not even know the number of component populations in the mixture distribution of the training sample. Also, it often occurs that some data will be missing, i.e. we will not always have all features measured at each of the events in the training sample or in the potential outlier. Finally, in a relatively new region, the training sample may actually contain a few unusual non-nuclear events (malfunctioning ripple-fired mining explosions, mine collapses, etc.) that do not belong to any component population of the appropriate mixture distribution.

In this report, we study the problem of detecting a nuclear event or other rare or unusual seismic signals in new or relatively unexplored regions for which training samples do not satisfy the assumptions imposed by Wang, et al. (1996). An outlier detection procedure which is appropriate for the setting described in the preceding paragraph is described in Section 2. In Section 3, we discuss the results of a simulation study designed to examine the new outlier detection procedures, and in Section 4 we apply the procedures to actual seismic data.

2. The Procedure

In this section, a procedure is presented for detecting outliers in a region for which limited information is available concerning the training sample of non-nuclear events. We extend the work of Wang et al. (1996), to develop an outlier-detection procedure that applies to the case in which no labeled training data are available in a region.

(*a*) *Data Types*

(*i*) *Known number of components and no missing data*

Here we assume that the training sample is known to contain a fixed number of event types (i.e. m in (1) is known.), and we assume that no data are missing. The event groups in the training sample represent the types of non-nuclear seismic activity in the region, e.g. earthquakes and mining blasts.

Let

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \Pi$$

denote the training sample of size n from the mixture population. In the notation of Redner and Walker (1984) the sample is of Type 1, i.e. it consists only of unlabeled observations. A new observation, $\boldsymbol{X}_{n+1}$, is obtained, and given the training sample we want to test the hypothesis

$$H_0 : \boldsymbol{X}_{n+1} \in \Pi$$

vs.

$$H_1 : \boldsymbol{X}_{n+1} \notin \Pi .$$

The classical likelihood ratio test statistic is the ratio of the maximized likelihood functions under H_0 and H_1. Under H_0 we assume that $\boldsymbol{X}_{n+1}$ is from the same mixture distribution as $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, that is the likelihood function under H_0 is

$$L_0(\theta) = \left(\prod_{s=1}^{n} f(\boldsymbol{X}_s;\boldsymbol{\theta})\right) f(\boldsymbol{X}_{n+1};\boldsymbol{\theta})\,.$$

If $h(\boldsymbol{x};\ \beta)$ denotes the density associated with the outlier population from which $\boldsymbol{X}_{n+1}$ is sampled, where β is an unknown parameter vector, then the likelihood function under H_1 is

$$L_1(\theta, \beta) = \left(\prod_{s=1}^{n} f(\boldsymbol{X}_s;\boldsymbol{\theta})\right) h(\boldsymbol{X}_{n+1};\ \beta).$$

Difficulties arise when maximizing L_1 since there is only a single observation from the outlier population. To overcome these difficulties and to acknowledge the fact that little information is known about the outlier population from which $\boldsymbol{X}_j$ is sampled, Wang et al. (1996) used a constant density $h(\boldsymbol{x}) \equiv c$ over its practical (finite) support. We use this approach here and let

$$\widetilde{L}_1(\theta) = \prod_{s=1}^{n} f(\boldsymbol{X}_s;\boldsymbol{\theta})$$

which is the likelihood based on the training sample $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ from the mixture, and we define a simple modified likelihood ratio test statistic

$$W = \frac{\sup_{\theta \in \Theta} L_0(\theta)}{\sup_{\theta \in \Theta} \widetilde{L}_1(\theta)}, \tag{2}$$

where Θ is the entire parameter space. It is easily seen that the departure of $\boldsymbol{X}_{n+1}$ from f will reduce $\sup_{\theta \in \Theta} L_0(\theta)$ making W smaller. Hence, the rejection region is of the form $W \leq W_\alpha$ for some W_α picked to provide a level α test. Since the null distribution of W has no known closed form, we use the nonparametric bootstrap to approximate it. Specifically, B bootstrap samples are obtained, $b = 1, \ldots, B$. Each bootstrap sample is

obtained by resampling from the training sample to obtain a sample of size $n+1$, and for each bootstrap sample the test statistic W_b^* is obtained where $b = 1, \ldots, B$. We then define W_α to be the (100α)th percentile of the W_b^*. Specifically, if $\alpha = j/(B+1)$, then W_α is the jth smallest value of $\{W_b^*\}_{b=1}^{B}$ (see McLachlan, 1987). For a discussion of the nonparametric bootstrap when some data are labeled, see Wang et al. (1996).

It should be noted that the maximum likelihood estimates involved in evaluating (2) are obtained using the EM algorithm (McLachlan, 1982 and Redner and Walker, 1984). This procedure is iterative in nature and requires initial values of the parameters estimates. Additionally, under the present scenario it is assumed that no initial estimates for the parameters, μ_i and Σ_i, of the component distributions or the mixing proportions, p_i, are available. Wang et al. (1996) assumed that a sufficient amount of labeled data are available to provide initial estimates. In our setting, this information is not assumed to be available and a clustering approach is used to group the data into distinct classes (see Appendix for details) from which initial estimates can be obtained. The initial estimates are then taken from the data assigned to each of the unknown classes.

(*ii*) *Number of components unknown and no missing data*

It will often be the case that the number of components, i.e. distinct classes in the population of the training sample will not be known. In this section, we propose a modification of the procedure in (*i*) which is appropriate when the number of components is unknown. We consider the use of Akaike's AIC (Akaike, 1974) for purposes of determining the number of components m in the mixture. The use of AIC has been considered in this setting by Sclove (1983), Bozdogan and Sclove (1984), Redner, Kitagawa, and Coberly(1984), and Gray, Woodward, and McCartor (1989). Specifically, for $m = 1, \ldots, M$ we calculate

$$\text{AIC}(m) = -2\ln(L_{max}(m)) + 2(\#\text{ of free parameters})$$

where $L_{max}(m)$ is the maximized likelihood of the training sample under the assumption that there are m components and M is a sufficiently large integer. $L_{max}(m)$ is obtained via the EM algorithm as discussed in the previous section. AIC imposes a penalty based on the number of parameters, and it should be noted that in the case in which the means, covariances and mixing proportions are unknown and the feature vector is of order d, then there are $md + md(d+1)/2 + m - 1$ free parameters so that even for relatively small d, the penalty increases rapidly as m increases. For each m, $m = 1, \ldots, M$, we use the clustering discussed previously to obtain m initial clusters to provide starting values for the EM algorithm from which $L_{max}(m)$ is obtained. AIC is calculated for $m = 1, \ldots, M$, and the number of components, m_{AIC}, associated with the minimum AIC is chosen. The test statistic for the data, W, is then calculated as in (2) based on m_{AIC} components.

To obtain the distribution of W, we again use the nonparametric bootstrap. The bootstrap samples are selected as before, and for the bth bootstrap sample we find $m^*_{\text{AIC}}(b)$ using AIC and calculate W^*_b on this basis. We then take W_α to be the (100α)th percentile of the W^*_b, $b = 1, \ldots, B$ as before.

It is well known that in general AIC does not provide a consistent estimator of the model order, and that the selected model order has the tendency to increase as sample size increases thus leading to overly complicated models. To compensate for this, in the simulations and data analysis in the next sections, and as an alternative to AIC we consider the use of BIC (Akaike, 1977) given by

$$\text{BIC} = -2\ln(L_{max}(m)) + \ln(n)(\,\#\text{ of free parameters})\,.$$

BIC imposes a more severe penalty than does AIC and in some cases provides a consistent estimator of the model order (e.g. Hannan, 1980).

It should be noted that in the case of mixtures, the regularity conditions for $-2\ln(L)$ to have its usual asymptotic chi-square distribution do not hold (see e.g. McLachlan and Basford, 1988). Despite the fact that AIC has been used successfully in mixture settings by several authors listed previously in this section, Titterington, Smith and Makow (1985) have shown that the theoretical justification of the use of AIC or BIC relies on basically these same regularity conditions.

In Section 3, we will investigate the use of AIC and BIC in order to determine their actual performance in the context of the bootstrap-based likelihood ratio outlier test considered here.

(*iii*) *Missing data*

In either (*i*) or (*ii*) it may be the case that some of the data are missing. Miller, Gray and Woodward (1993) studied outlier testing in the setting in which the training sample is from a multivariate (non-mixture) population when some data are missing. They considered the use of the EM algorithm versus simple mean replacement for dealing with missing data, and their findings were that the performance of mean replacement (at least for no more than 20% missing) was comparable with the full EM algorithm at a fraction of the computation requirements. Based on these findings, we considered the use of a mean replacement strategy for dealing with missing data in the mixture setting considered here. If $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})'$ denotes the ith observation in the training sample and if, for example, x_{i2} is missing, then in the non-mixture setting, mean replacement consists of simply replacing this missing observation with the sample mean of feature 2 across all sample values for which this feature was actually observed. In the mixture setting, however, we would want to replace the missing x_{i2} by the sample mean of existing observations in the component to which observation x_{i2} belongs. When the

training sample data are labeled, then this procedure is easily accomplished. However, when data are not labeled, an initial clustering is required in order to ascertain the component to which x_{i2} most likely belongs. Obviously, this initial clustering must take into account the missing data in such a way that the distance between two observations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be calculated even if one or both of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ has missing data on some features. Specifically, we use a procedure suggested by Dixon (1979) to calculate the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as

$$d(i,j) = \frac{d}{d-d_0} \sum_{k=1}^{d} d_k^2$$

where

$$d_k = \begin{cases} 0 & \text{if } x_{ik} \text{ or } x_{jk} \text{ is missing} \\ x_{ik} - x_{jk} & \text{otherwise} \end{cases}$$

and where d_0 is the number of features missing in $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. It can be seen that $d(i,j)$ is simply the squared Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ whenever there are no missing features in $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. In order for this to be a reasonable procedure, it is important to first standardize the data as mentioned in the appendix so that the (non-missing) data on each feature have unit sample variance.

For a given number of components we perform the cluster analysis based on the available data using the metric $d(i,j)$. Once the clusters are established using the k-means algorithm, we replace a missing feature in a data value with the sample mean of existing observations of that feature in the cluster to which the data value belongs. Once the mean replacement is accomplished, then the likelihood ratio calculations can be performed using the newly created "completed" data set. If the number of components is known, then W is calculated from the "completed" data set. Note that the mean replacement depends on the number of components assumed, so in the case in which the

number of components is unknown and AIC or BIC is used to estimate it, a separate mean replacement will be required for each m, $m = 1, \ldots, M$.

In order to ascertain the distribution of W, the bootstrap is applied as before. It should be noted that the resampling is done from the original raw data, and thus some of the data in the bootstrap samples may be missing if this were the case for the original data. In the case of known number of components, m, the bootstrap is analogous to that used in (i). That is, for each bootstrap sample, m clusters are formed using the $d(i, j)$ metric, the corresponding mean replacement is done, and W_b^* is calculated using the "completed" data. For the case in which the number of components is not known, the AIC (or BIC) will be obtained for each bootstrap sample as in (ii).

(b) *Cleaning the Training Sample*

While it is assumed that the training data contain no nuclear events, the procedure we propose includes an examination of the training sample for unusual events (i.e. events which are in fact outliers themselves such as mine collapses) that should be removed before the training sample is used for testing new, and possibly nuclear, events.

After the initial estimates are obtained, each point in the training sample is considered individually by using the other $n - 1$ points as a pseudo training sample. The modified likelihood ratio test developed in Wang et al. (1996) is used to test each point and determine the probability that each point belongs to the assumed mixture population. Any point with a significant result (very small probability of inclusion in the mixture - say 0.01 level) at this phase is labeled as an outlier and is removed from the training data set. After checking each point, the remaining points are then used as a "clean" training data set for testing future events as potential outliers from the mixture population.

3. Simulations

In this section the effect of unlabeled data, unknown number of components and missing data on the detection probability of the outlier test based on W is examined using

a simulation study based on the procedures described in Section 2. In the simulations of this section, the training data are from a mixture distribution as in (1) with $m = 2$ and where the component distributions are each bivariate normal. Specifically, components 1 and 2 are distributed

$$N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0.5\\0.5 & 1\end{pmatrix}\right)$$

and (3)

$$N\left(\begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}1 & -0.5\\-0.5 & 1\end{pmatrix}\right)$$

respectively with $p_1 = p_2 = 0.5$. For the simulation, training samples of size $n = 60$ are generated from this mixture population, and outliers are generated from the populations

$$N\left(\begin{pmatrix}1+k-5\\1-k+5\end{pmatrix}, \begin{pmatrix}1 & 0.5\\0.5 & 1\end{pmatrix}\right), \quad (4)$$

where $k = 1, ..., 9$. In Figure 1, we show the contours of the mixture components along with the outlier means $(1 + k - 5,\ 1 - k + 5)'$, $k = 1, ..., 9$. In this figure, we also show the contour of the outlier population for the case $k = 2$, i.e. the mean is $(-2,\ 4)'$. All tests are based on an $\alpha = 0.05$ nominal significance level.

In Table 1, we show the results for the case of known components and for various degrees of labeling. The table shows detection probability results for the case in which all training sample observations are labeled using the technique based on W. The estimates shown are the proportion of the 1000 replications for which an outlier was detected. In general, the lack of labeling information from the class labels of the training data leads to no detectable decrease in detection probability as would be expected. This suggests that in these types of settings, although lack of labeling may degrade our

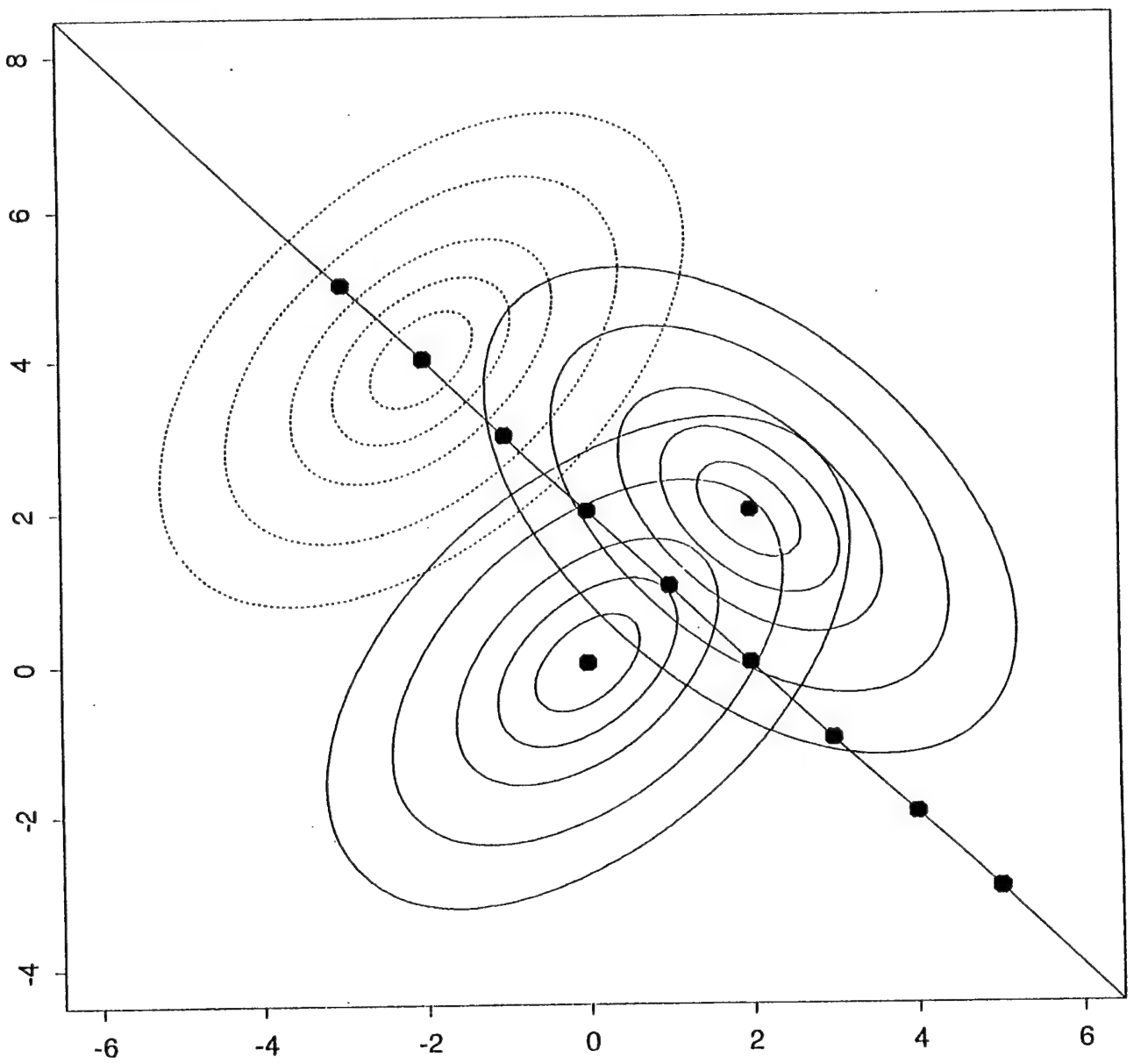

Figure 1: Mixture distributions for Example 2*a* showing means of outlier distributions in the simulations

estimates of the components of the mixture, it does not seem to have a dramatic effect on the estimated mixture distribution itself.

Next, we consider the case in which the number of components is not known. We generated 100 samples of size $n = 60$ from the mixture distribution in (3), and in Table 2 we show the number of clusters identified by AIC and BIC respectively where the maximum allowable number of clusters is taken to be $M = 4$. In the table, it can be seen that AIC does the better job of correctly identifying $m = 2$ as the number of components while it rarely underestimates m and overestimates it 36% of the time. On the other hand, BIC selected $m = 2$ only 49% of the time, underestimated m on 51% of the cases and never overestimated it. These results are consistent with the discussion of AIC and BIC in Section 2. In order to examine the implications on the outlier test of not knowing the number of components, in Table 3 we show power corresponding to that shown in Table 1 when m is unknown but estimated by either AIC or BIC. In these simulations, we generated 100 replicates of size $n = 60$ from the two-component distribution in (3) for which none of the observations were labeled and for which the number of components was assumed to be unknown in the analysis stage. It can be seen in Table 3 that there is some loss in power when compared to the known component, 100% unlabeled case in Table 1. However, the powers are not dramatically smaller. The power results using AIC and BIC are very similar. Additionally, the observed significance levels are slightly higher than the nominal $\alpha = 0.05$ level.

In order to examine the effect of missing observations, we simulate 100 samples of size $n = 60$ from the mixture model in (3) where the number of components is assumed to be known and where a percentage of the observations are taken to be missing.

Table 1. Significance levels and power of outlier test with some data unlabeled and number of components assumed to be known.

$n = 60$, 1000 replications

Training sample population : 2-component mixture in (3) with $p_1 = p_2 = 0.5$
Outlier populations: specified by $k = 1, \ldots, 9$ as in (4)

		k								
% Unlabeled	**Sig. Level**	1	2	3	4	5	6	7	8	9
0	.062	1.00	.979	.799	.260	.031	.249	.756	.983	1.00
10	.074	1.00	.979	.796	.282	.037	.291	.768	.977	1.00
25	.068	1.00	.981	.783	.254	.032	.264	.767	.985	1.00
50	.061	1.00	.980	.790	.255	.032	.249	.758	.985	1.00
75*	.061	.999	.980	.767	.232	.030	.237	.746	.972	1.00
100	.066	1.00	.986	.761	.261	.032	.251	.747	.981	.999
S.E.	.007	0.015								

* Several of these (18) lacked sufficient data in a group for starting values.

Table 2. AIC and BIC selections of the number of components expressed as proportion of 100 samples of size 60 from the 2-component mixture model in (3) with $p_1 = p_2 = 0.5$

	Number of components selected			
	1	2	3	4
AIC	.03	.61	.26	.10
BIC	.51	.49	0.0	0.0

Table 3. Significance levels and power of outlier test with all data unlabeled and the number of components assumed unknown.

$n = 60$, 100 replications

Training sample population : 2-component mixture in (3) with $p_1 = p_2 = 0.5$
Outlier populations: specified by $k = 1, \ldots, 9$ as in (4)

		k								
Criterion	**Sig. Level**	1	2	3	4	5	6	7	8	9
AIC	.08	1.0	.98	.70	.27	.04	.21	.70	.96	1.0
BIC	.09	1.0	.98	.75	.23	.01	.17	.67	.97	1.0
S.E.	.02	.05								

Table 4. Significance levels and power of outlier test with all data unlabeled, number of components assumed known, and some data missing.

$n = 60$, 100 replications

Training sample population : 2-component mixture in (3) with $p_1 = p_2 = 0.5$
Outlier populations: specified by $k = 2, \ldots, 5$ as in (4)

		k			
% missing	**Sig. Level**	2	3	4	5
10	.09	.99	.75	.27	.02
25	.12	.99	.73	.32	.04
50	.19	1.0	.83	.35	.04
S.E.	.02	.05			

In Table 4 we see that the observed significance levels seem to be inflated above the $\alpha = 0.05$ level, especially for greater than 10% missing. The effect of the current mean-replacement is to excessively reduce the within cluster variability. These results indicate that if a substantial percentage of data will be missing, then improved procedures for handling the missing data must be developed and tested. Possibilities include use of the EM algorithm as was done by Miller et al. (1993) or a replacement procedure similar to that used above but for which the missing value is replaced by something other than the midpoint of the cluster.

4. Example using Seismic Data

The data for this example are based on an analysis of earthquakes and mining explosions from the Vogtland region near the Czech-German border by Burlacu and Herrin (1996). These data were taken from the ground truth database compiled by Grant, et al. (1993). These measurements are new to the seismic community and involve fitting a third order autoregressive process to the S wave. The power spectral density is estimated and the strength and frequencies of the real and complex poles are calculated. These are useful features since distributed surface explosions (i.e. ripple-fired mining blasts) tend to be lower frequency with a sharper spectrum (strong pole) and earthquakes tend to have higher frequency and a more distributed spectrum (weak pole). These features are incorporated into a promising screening process to identify mining blasts. In the analysis here, the complex frequency and pole strength associated with an AR(3) fit to the data are used as feature variables.

Table 5 contains information on the events used in this study and Figure 2 shows a scatter plot of the complex frequency and pole for each event (plotting characters indicate event number). Note that event number 25 is listed in the ground truth data base as an explosion, although some controversy has surrounded this event. For this example, the ground truth information is not used. Rather, the source for each event is assumed to

Table 5. Information on the 27 Vogtland events

Event #	Date	Lat(N)	Long(E)	Depth	M_1	Y(kg)	Or.time	Q/X
1	031191	50.207	12.685	0	1.98	3,265	12:03:24	EX
2	032191	50.207	12.685	0	2.05	3,982	12:04:15	EX
3	032291	50.207	12.685	0	2.03	2,835	12:33:25	EX
4	032391	50.207	12.685	0	1.99	2,025	12:00:56	EX
5	032491	50.296	12.225	12.9	2.18	-	05:05:04	EQ
6	032491	50.279	12.228	12.9	1.50	-	05:35:21	EQ
8	032491	50.278	12.220	12.4	1.65	-	09:38:33	EQ
9	032491	50.294	12.223	12.7	2.07	-	14:33:28	EQ
10	032491	50.293	12.224	12.5	1.80	-	15:00:45	EQ
11	032491	50.293	12.224	9	1.73	-	15:41:04	EQ
12	032591	50.298	12.222	12.9	2.37	-	14:54:14	EQ
13	032591	50.292	12.213	12.4	1.54	-	22:31:46	EQ
15	050291	50.207	12.713	0	1.93	3,575	11:06:10	EX
19	051991	50.360	12.371	0	2.06	-	03:22:10	EQ
20	052391	50.207	12.713	0	2.12	3,135	11:01:05	EX
21	052591	50.207	12.713	0	2.13	3,135	11:01:29	EX
23	052891	50.207	12.685	0	2.01	3,575	11:03:51	EX
24	062091	50.207	12.685	0	1.98	1,998	11:01:17	EX
25	062091	50.293	12.803	0	1.80	-	11:45:35	EX
26	062291	50.207	12.685	0	2.15	2,886	10:58:34	EX
27	062791	50.207	12.685	0	1.93	3,515	11:04:40	EX

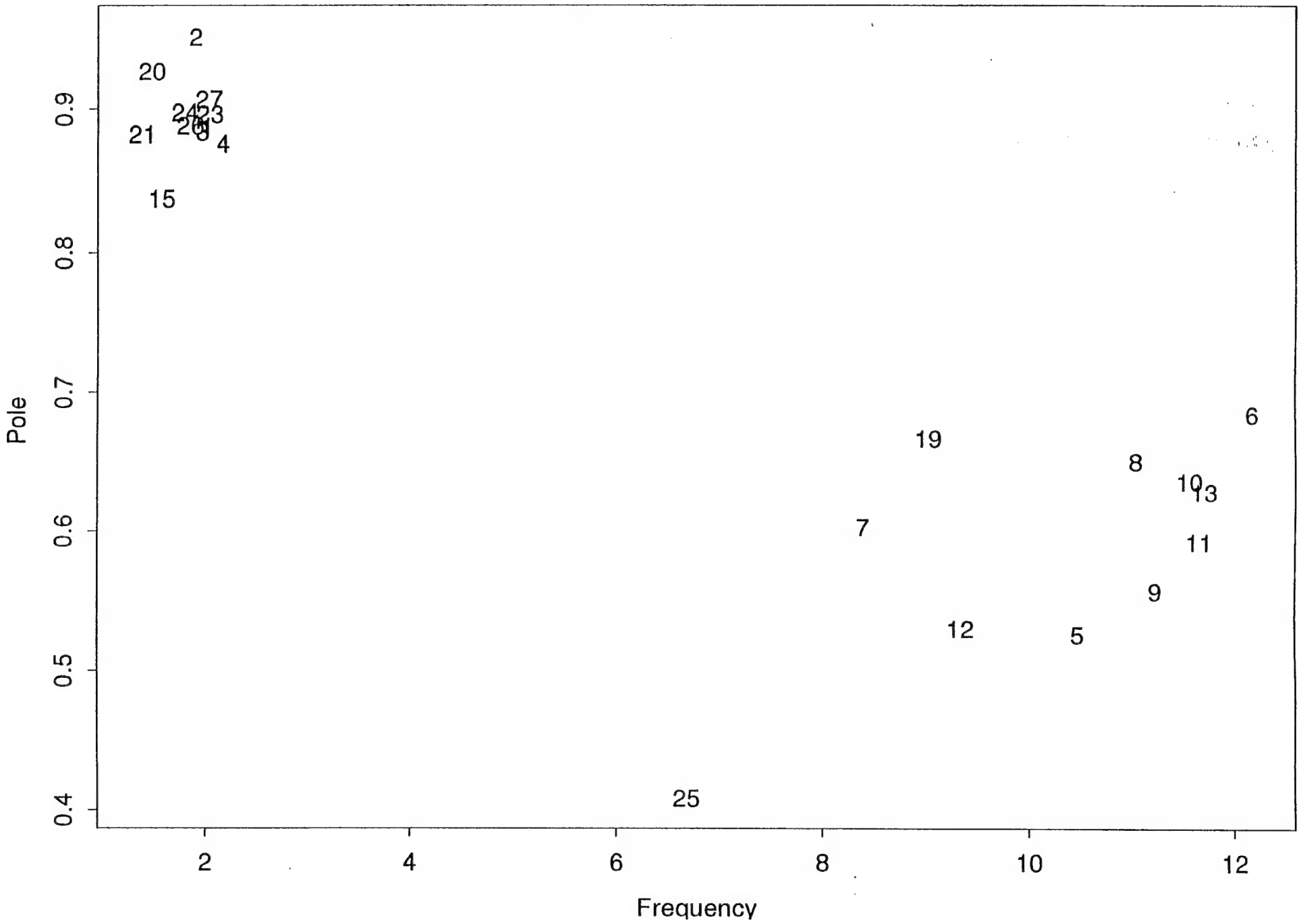

Figure 2: Scatterplot of Frequency vs. Pole

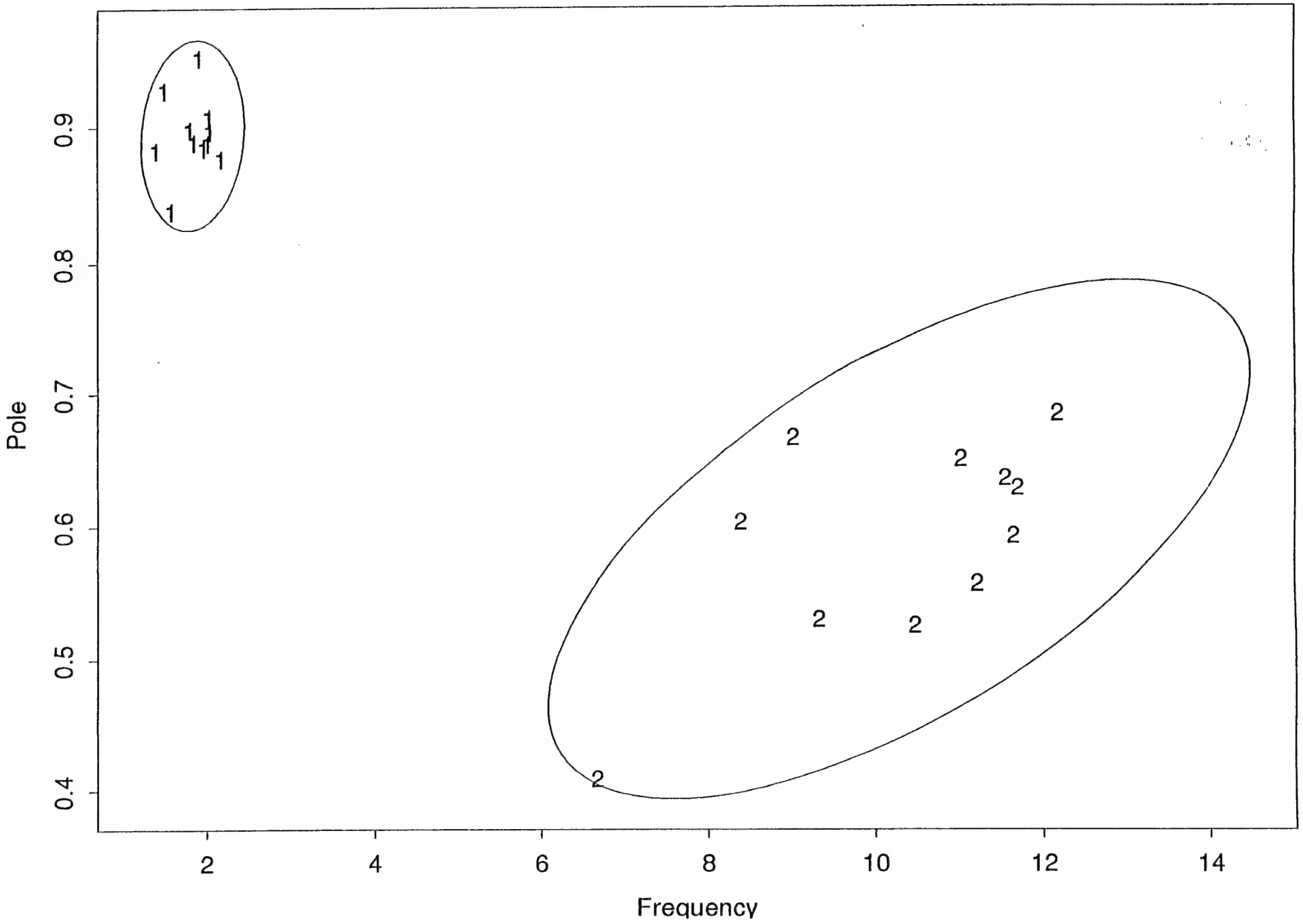

Figure 3: Results of Cluster Analysis

be unknown, although it is assumed that the data set is composed of observations from two sources (earthquake and mining blast). Finally, it is also assumed that no nuclear events are present in the training sample.

Figure 3 shows the result of the cluster analysis. The members of each cluster are indicated on the plot (as a "1" or a "2") as well as a 95% contour for each component normal distribution using the parameters estimated from the results of the cluster analysis. Note that the labeling of clusters is arbitrary and does not indicate the source of the event. These data show a clear separation between the groups. Hence, only one iteration of the cluster analysis is necessary, i.e. no observations in the training sample were determined to be outliers.

Figure 4 shows the results of the leave-one-out testing procedure. Plotted are the p-values for being in the mixture associated with each frequency and pole pair (plotting characters indicate p-value). Note that only event 25 shows a significant result (p-value < 0.01), which leads to the conclusion that event 25 is an outlier to the mixture distribution of earthquakes and explosions. Results for all other points support their membership in the mixture and are consistent with the ground truth information.

New events in this region should now be tested using this "clean" data. Figure 5 shows contours representing effective rejection regions ($\alpha = 0.1, 0.05, 0.01$) based on this training sample. Note that these regions mirror the shape of the distributions suggested by the data.

5. Concluding Remarks

In this report, we show that outlier detection based on a mixture training sample of totally unlabeled data can be successfully accomplished even when the number of components is not known and when some data are missing.

The focus of the current report is on the case in which all data are unlabeled whereas Wang et al. (1996) assumed that some (or all) of the data were labeled with

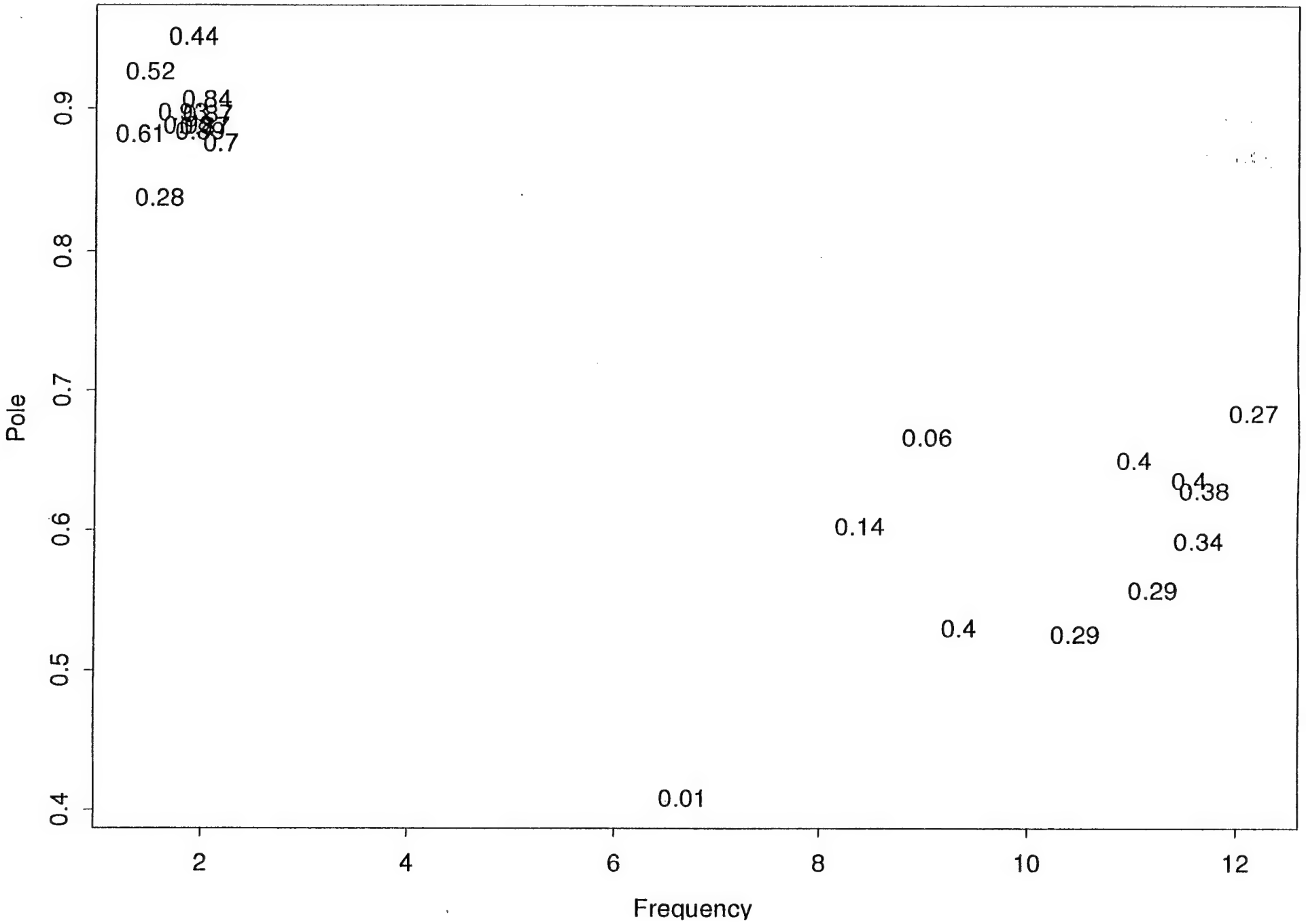

Figure 4: Results of Leave-One-Out Analysis

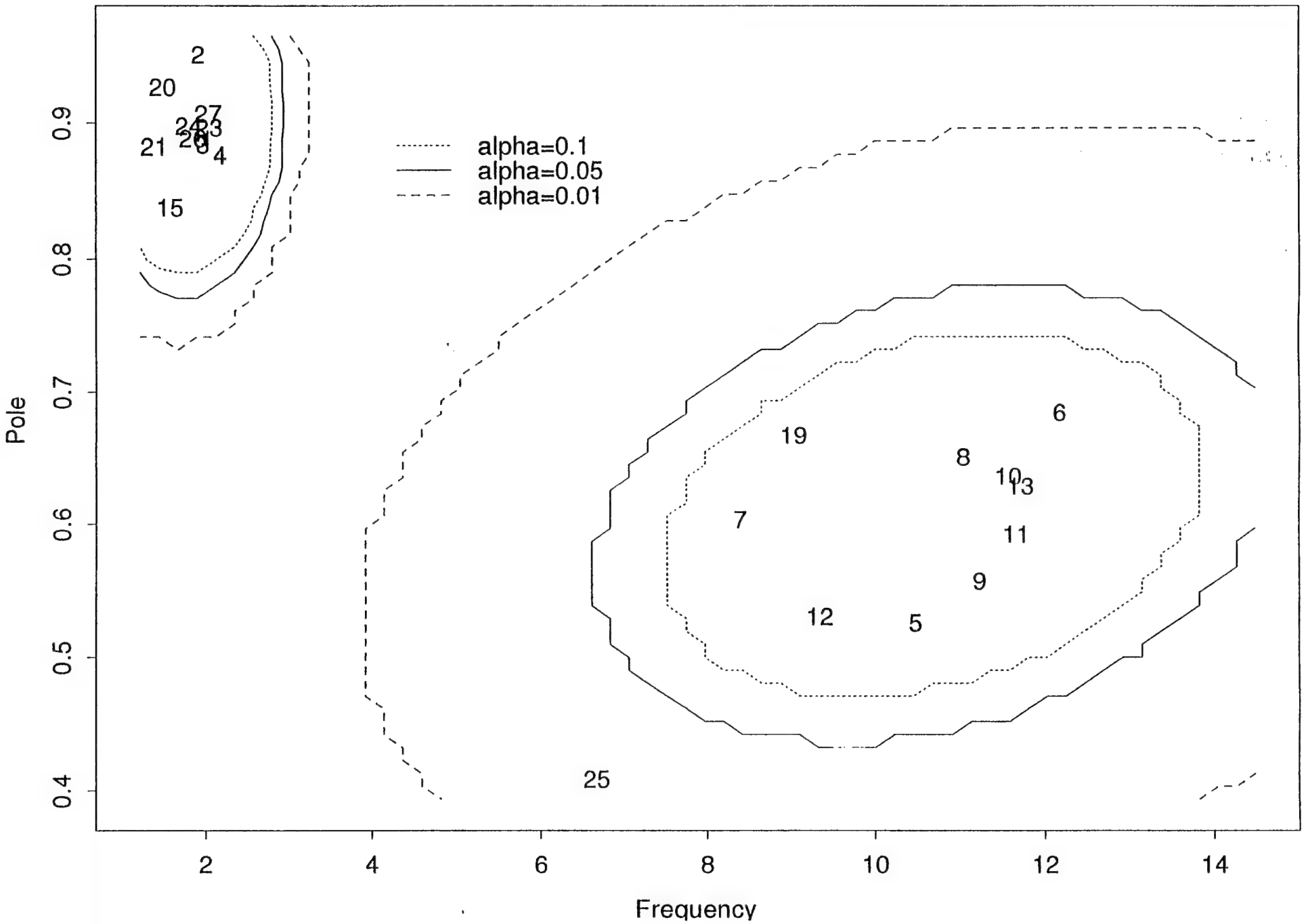

Figure 5: Rejection Regions

regard to their component membership. However, missing data may still be a problem in the case in which some or all labeling is known. If all data are labeled, then it is clear that mean replacement (or an alternative algorithm) for an observation in component i would be based on existing data from the sampled values in the ith component. When some data are labeled and some are unlabeled, then missing data in a labeled observation would be handled as just mentioned for the fully-labeled case. Each unlabeled observation is assigned to the component to which it is closest (using $d(i, j)$). Once component membership is thus established, a missing value for an unlabeled observation is replaced using the labeled data in that component. It should be noted that in this discussion we have assumed that if some or all of the data are labeled, then the number of components is known.

It is desirable to assign labels to events in the training sample after the clustering and estimation of component parameters is accomplished. We first consider the case in which the training data consist of two components. Each point in the training sample is tested as an outlier from each of the two training sample components and corresponding p-values obtained are associated with each component. Based on these p-values, each training sample member would be assigned a component membership or will be left unassigned when membership is not clear as defined by some predetermined p-value. Use of tests based on a focused critical region can be used to increase our ability to assign component membership based on the position of the training sample value being tested with respect to the locations of the corresponding component centroids. When the distribution of the training sample has more than two components, the testing can be based on considering the components two at a time. Actual "naming" of components can be done by an analyst, or by a defined statistic and/or auxiliary variables.

Appendix: Clustering

Clustering is the process of grouping similar objects on the basis of characteristics of the objects. For a general treatment of the subject, see, for example, Hartigan (1975), Jain and Dubes (1988), and multivariate analysis texts such as Seber (1984).

Two basic types of clustering algorithms are used here. Before implementing either of these clustering techniques, we will first standardize the data so that the (non-missing) data on each feature have zero sample mean and unit sample variance. The first clustering technique considered is hierarchical clustering, which is an iterative technique involving the grouping of smaller clusters into larger ones until the desired number of clusters has been achieved. The second type partitions objects into non-overlapping groups by setting the number of clusters, choosing initial locations of the clusters, and then assigning points to one of the groups according to some pre-specified criterion. The k-means approach of Hartigan (1975) is an example of this second type of clustering. We take a two-stage approach to clustering. First, a hierarchical approach is used to obtain initial parameter estimates of the clustering. Then, in some cases, a procedure similar in nature to the k-means approach is used to refine the parameter estimates.

The hierarchical clustering algorithm begins by considering each of the n data points as an individual cluster. Then, the two points nearest to each other are combined to form $n - 1$ clusters. The procedure continues by combining or fusing the two clusters that are the most similar at each iteration. Similarity is a distance measure that can be calculated in a variety of ways. We use the nearest centroid method, which measures similarity as the distance between the centroids or means of the points in each cluster, due to the fact that it is more robust than single and complete linkage measures.

The hierarchical approach can be effected by extreme observations, particularly in situations where there is some overlap in the distributions of the data. For example,

consider the data presented in Figure 6. Here, 60 observations are generated from two bivariate normal distributions with equal probability. These distributions are indicated by the ellipses given in the first plot of Figure 6. The second plot shows the results of the hierarchical approach. Note the small cluster near (2,4) whose observations are labeled on the plot as a '2'. At the last step, the number of clusters is reduced from three to the required two. Since the distance from this smaller cluster to the other two is greater than that between the other two, it remains an individual cluster and the others are joined together. Clearly the parameter estimates from such clusters would be biased due to this probable poor clustering.

To prevent such problems, the second phase of the clustering is applied. Namely, each object is checked to see if it has been clustered in a reasonable way. If not, the object is reassigned to a more appropriate cluster. Many measures have been considered for determining the appropriateness of the cluster labels. However, these are often quite complex and difficult to compute. In this work, we use a precursor to the k-means approach suggested by Forgy (1965). The distance between each object and the cluster centers is calculated. If the object is not assigned to the cluster to which it is closest, then it is reassigned to that cluster. After all objects are checked in this fashion, parameter estimates are updated and the procedure is repeated until no objects are reassigned.

This reassignment tends to produce clusters of roughly equal size, so care must be used when the expected proportions of observations in each cluster are not equal. However, the parameter estimates are used merely as starting values for the EM algorithm which is fairly robust to initial parameters estimates. As a final note, if the clusters are sufficiently separated, then the reassignment will be unnecessary.

Finally, it is possible that the clustering algorithm will return one or more clusters that are considerably smaller than is reasonable, even with the reassignment. In our implementation, these small clusters are temporarily set aside, and the remaining data are used to form clusters and estimate initial parameters.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* 19, 716-723.

Akaike, H. (1977), "On Entropy Maximization Principle," *Applications of Statistics*, Ed. P.R. Krishnaiah, Amsterdam: North-Holland, 27-41.

Berlacu, R. and Herrin, E. (1996), "Discrimination between Earthquakes and Ripple-Fired Explosions Based on a Parametric Spectral Approach," Private communication.

Bozdogan, H. and Sclove, S.L. (1984), "Multi-sample Cluster Analysis U sing Akaike's Information Criterion," *Annals of the Institute of Statistical Mathematics* 36, 163-180.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (with discussion), *Journal of the Royal Statistical Society B39,* 1-38.

Dixon, J.K. (1979), "Pattern Recognition with Partly Missing Data," *IEEE Transactions on Systems, Man and Cybernetics, SMC* 9, 617-621.

Forgy, E.W. (1965), "Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications," *Biometrics*, **21**, 768-769.

Grant. L., Coyne, J., and Ryall, F. (1993), "CSS Ground Truth Database: Version 1 Handbook," *Scientific Report*, C93-05, Center for Seismic Studies, Arlington, VA.

Hannan, E.J. (1980), "The Estimation of the Order of an ARMA Process," *Annals of Statistics* 8, 1071-1081.

Hartigan, J.A. (1975), *Clustering Algorithms*, John Wiley & Sons, New York.

McLachlan, G.J. (1982), "The Classification and Mixture Maximum Likelihood Approaches to Clustering," in *Handbook of Statistics, Vol. 2,* P.R. Krishnaiah and L.Kanal (Eds.). Amsterdam: North-Holland, 199-208.

McLachlan, G.J. (1987), "On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture," *Applied Statistics*, 36, 318-324.

McLachlan, G.J. and Basford, K.E. (1988), *Mixture Models: Inference and Applications to Clustering*, New York: Marcel Dekker, Inc.

Miller, J.W., Gray, H.L., and Woodward, W.A. (1993),"Discriminant Analysis and Outlier Testing when Data are Missing," Semi-Annual Technical Report Four, Advanced Research Projeacts Agency, Nuclear Monitoring Research Office, August 19, 1993.

Redner, R.A., Kitagawa, G. and Coberly, W.A. (1981), "The Akaike Information Criterion and its Application to Mixture Proportion Estimation," NASA Technical Report No. SR-TI-04207.

Redner, R.A. and Walker, H.F. (1984), "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, **26**, 195-239.

Sclove, S.L. (1993), "Application of the Conditional Population-Mixture Model to Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-5, 428-433.

Seber, G.A.F. (1984), *Multivariate Observations*, New York: John Wiley & Sons.

Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985), *Statistical Analysis of Finite Mixture Distributions,* New York: John Wiley & Sons.

Wang, S., Woodward, W.A., Gray, H.L., Wiechecki, S., and Sain, S.R. (1996), "A New Test for Outlier Detection from a Multivariate Mixture Distribution," *Journal of Computational and Graphical Statistics*, Submitted.

Appendix B

A NEW LOOK AT OUTLIER TESTS WITH MULTIPLE STATIONS

Wayne A. Woodward*, Suojin Wang**, H.L. Gray*, and William H. Frawley*
* Southern Methodist University
**Texas A&M University

ABSTRACT

We address the problem of using regional seismic data to distinguish between nuclear events and events such as earthquakes, mining explosions when events are observed at several stations within a region. We use a bootstrap-based outlier testing approach to test whether a suspicious event should be considered to be an outlier from the population of the training sample. Because there may be several stations with several features measured at each station, straightforward use of all data at all stations may result in variance/covariance matrices of large order, e.g. as large as 80×80. Thus, it is important to develop data compression procedures that, for example, combine results for a given feature across stations. The results in the current paper extend the results of Fisk, Gray and McCartor (1995) and Gray, Woodward, and Yücel (1995). In this report, we develop a new set of weights for combining station information that are shown to perform better in simulations than the minimum variance weights considered by Fisk et al. (1995) and Gray et al. (1995). A "double-weighting" approach is also considered. We briefly consider the case in which the population of the training sample is considered to have a mixture distribution which allows for the existence of more than one type of non-nuclear event in a region. i.e. earthquakes and mining explosions.

1. Introduction

The problem of observing seismic events for the purpose of distinguishing between nuclear events and events such as earthquakes, mining explosions, etc. has been studied by several authors. The usual scenario in which this problem has been considered is to consider the existence of a training sample of non-nuclear events in a region and to test new and possibly suspicious events as possible outliers from the population of the training sample. Baek, Gray, McCartor, and Woodward (1992) use a bootstrap likelihood ratio test to determine whether an event should be considered to be an outlier from a single multivariate population where measurements were made at a single station. Miller, Gray, and Woodward (1993) extend this test to the case in which some data are missing. Fisk, Gray and McCartor (1995) and Gray, Woodward, and Yücel (1995) extended these results to cover the case in which readings are obtained at multiple stations. Wang, Woodward, Gray, Wiechecki, and Sain (1996) consider the problem of testing an event as an outlier from a mixture population which represents the realistic scenario in which there may be more than one type of non-nuclear event in a region. The work of Wang et al. (1996) allows for the training sample to represent a sample from, for example, earthquakes and mining blasts; but this was based on data from a single station.

In this report, we consider the use of outlier tests when readings are obtained from multiple stations. We re-examine the scenario considered by Fisk et al. (1995) and Gray et al. (1995), and introduce a modification of the likelihood ratio approach employed in the papers cited which is more suitable for multistation data when the outlier population is considered to be a mixture of event types.

2. Review of Previous Multistation Results

In this section, we assume that d features are measured on n events detected at m stations. We let X_{jki} denote the measurement of the kth feature for the ith event in the training sample measured the jth station. That is, for the kth feature, we have the following training data:

Station 1	...	**Station m**
X_{1k1}	...	X_{mk1}
$\vdots$		$\vdots$
X_{1kn}	...	X_{mkn}

We use the notation $\boldsymbol{X_{ki}} = (X_{1ki}, \ldots, X_{mki})'$ to denote the m station readings for the kth feature and ith event, $\overline{X}_{jk}$ to denote the average of the n events measured at station j and feature k, and $\overline{\boldsymbol{X}}_{\boldsymbol{k}} = (\overline{X}_{1k}, \ldots, \overline{X}_{mk})'$ to denote the vector of these averages evaluated at each of the m stations. The m station readings for the potential outlier at the kth variable are denoted by $\boldsymbol{U_k} = (U_{1k}, \ldots, U_{mk})'$. For the present we consider the case in which the population of the training sample is a single (non-mixture) multivariate population.

Several approaches were considered by Fisk et al. (1995) and by Gray et al. (1995) for analyzing multistation data, and these will be briefly discussed here.

(a) Full Vector Approach

A full-vector approach was considered in which the d features at each of the m stations are considered as a single vector consisting of md variables. In this approach the observation vector for the ith event in the training sample is considered as an $md \times 1$ vector of the form

$$\boldsymbol{X_i} = (X_{11i}, X_{12i}, \ldots, X_{1di}, X_{21i}, X_{22i}, \ldots, X_{2di}, \ldots, X_{m1i}, X_{m2i}, \ldots, X_{mdi})' .$$

A new observation to be tested as an outlier is then a similarly configured $md \times 1$ vector denoted by

$$\boldsymbol{U} = (U_{11}, U_{12}, \ldots, U_{1d}, \ldots, U_{m1}, U_{m2}, \ldots, U_{md})' .$$

In the full-vector approach considered by Fisk et al. (1995) and Gray et al. (1995), the training sample $\{\boldsymbol{X_i}\}_{i=1}^{n}$ was considered to be from the density function $f(\cdot\,;\, \boldsymbol{\mu_1},\, \boldsymbol{\Sigma})$, where

$$f(\boldsymbol{x};\, \boldsymbol{\mu_1},\, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{mp}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\{ -\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_1}),$$

i.e. they assumed that the feature variables have a multivariate normal distribution. Similarly, the new $\boldsymbol{U}$ was assumed to have probability density $f(\cdot\,;\, \boldsymbol{\mu_2}, \boldsymbol{\Sigma})$. Baek, et al. (1992) classify $\boldsymbol{U}$ by testing the hypotheses

$$H_0 :\ \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$$
$$H_1 :\ \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}.$$

The generalized likelihood ratio is given by

$$\lambda = \frac{\sup_{\{\boldsymbol{\theta} \in \Omega_0\}} L(\boldsymbol{\theta};\, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{U})}{\sup_{\{\boldsymbol{\theta} \in \Omega\}} L(\boldsymbol{\theta};\, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{U})}$$

$$= \frac{L(\widehat{\boldsymbol{\theta}}_0;\, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{U})}{L(\widehat{\boldsymbol{\theta}};\, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{U})} \qquad (1)$$

where $\widehat{\boldsymbol{\theta}}_0$ is the Maximum Likelihood Estimate (MLE) of $\boldsymbol{\theta}$ under the restriction that H_0 is true, and $\widehat{\boldsymbol{\theta}} = \{\widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\mu}}_2, \widehat{\boldsymbol{\Sigma}}\}$ where $\widehat{\boldsymbol{\mu}}_1$ and $\widehat{\boldsymbol{\Sigma}}$ are the MLE's of $\boldsymbol{\mu_1}$and $\boldsymbol{\Sigma}$ based on $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ and $\widehat{\boldsymbol{\mu}}_2 = \boldsymbol{U}$. It intuitively follows that small values of λ provide evidence against H_0, and thus the generalized likelihood ratio test is to reject H_0 if $\lambda \leq \lambda(\alpha)$, where $\lambda(\alpha)$ is chosen to provide

a size α test. In some cases it will be necessary to approximate the critical value $\lambda(\alpha)$ using bootstrap techniques. In the setting considered here, i.e. the populations are multivariate normal, the above likelihood ratio test is equivalent to Hotelling's T^2 and, in fact, T^2 is proportional to $1/\lambda$. Recall, in the general case in which $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_{n_1}$ is a random sample from a multivariate normal population having density $f(\cdot\,; \boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{n_2}$ is an independent random sample from a multivariate normal population having density $f(\cdot\,; \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, then Hotelling's T^2 is given by

$$T^2 = \left(\tfrac{1}{n_1} + \tfrac{1}{n_2}\right)^{-1} (\bar{\boldsymbol{W}} - \bar{\boldsymbol{Z}})' \boldsymbol{S}_p^{-1} (\bar{\boldsymbol{W}} - \bar{\boldsymbol{Z}})$$

where S_p is the usual pooled estimate of the variance/covariance matrix

$$\boldsymbol{S}_p = \frac{(n_1-1)\sum_{i=1}^{n_1} (\boldsymbol{W}_i - \bar{\boldsymbol{W}})(\boldsymbol{W}_i - \bar{\boldsymbol{W}})' + (n_2-1)\sum_{i=1}^{n_2} (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}})(\boldsymbol{Z}_i - \bar{\boldsymbol{Z}})'}{(n_1+n_2-2)}.$$

where $\boldsymbol{W}_i = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_{n_1})'$, etc. In our setting, i.e. $n_1 = n$ and $n_2 = 1$, this becomes

$$\begin{aligned} T^2 &= \left(\tfrac{1}{n} + 1\right)^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{U})' \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{U}) \\ &= \left(\tfrac{n}{n+1}\right) (\bar{\boldsymbol{X}} - \boldsymbol{U})' \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{U}). \end{aligned} \tag{2}$$

Here, $\boldsymbol{S}$ must be calculated entirely on the basis of the training sample since $n_2 = 1$, and in this case $\boldsymbol{S}$ is simply the sample variance/covariance matrix based on the training sample data, i.e.

$$\boldsymbol{S} = \frac{\sum_{i=1}^{n_1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'}{(n_1-1)}.$$

(b) Minimum Variance Weighting

In the full vector approach, discussed in (a) no attempt is made to account for the fact that the same d variables are being measured at the m stations. Fisk et al. (1995) and Gray et al. (1995) considered a minimum variance weighting in an attempt to combine features across stations to reduce the dimensionality of the problem by taking advantage of the correlation structure between stations. In particular, they constructed a new "feature", Y_k, associated with feature k which is a linear combination of feature k at each of the m stations. We use the notation

$$Y_{ki} = \sum_{j=1}^{m} w_{jk} X_{jki}, \qquad i = 1, \ldots, n.$$

The weights $\boldsymbol{w_k} = (w_{1k}, \ldots, w_{mk})'$ were chosen to be those which minimize the variance of Y_{ki} subject to the constraint that the weights sum to one. Theoretically, the weights are given by $\boldsymbol{w_k} = \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\mathbf{1}/\mathbf{1}'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\mathbf{1}$ where $\mathbf{1}' = (1, 1, \ldots, 1)$ and $\boldsymbol{\Sigma_k}$ is the variance-covariance matrix of $\boldsymbol{X_{ki}}$. In practice $\boldsymbol{\Sigma_k}$ will not be known and will be estimated by the usual sample variance-covariance matrix, $\boldsymbol{S}_k$, based on events $i = 1, \ldots, n$. Thus, the weights are

$$\widehat{\boldsymbol{w}}_{\boldsymbol{k}} = \boldsymbol{S}_{\boldsymbol{k}}^{-1}\mathbf{1}/\mathbf{1}'\boldsymbol{S}_{\boldsymbol{k}}^{-1}\mathbf{1}. \tag{3}$$

This procedure creates the new d-dimensional vector $\boldsymbol{Y_i} = (Y_{1i}, \ldots, Y_{di})'$, $i = 1, \ldots, n$, and the potential outlier is weighted using these same weights, i.e.

$$V_k = \sum_{j=1}^{m} \widehat{w}_{jk} U_{jk}.$$

This weighting reduces the dimension from md variables to d variables which may be important when the number of features and stations becomes large. The outlier detection is then based on a likelihood ratio as before but calculated using only the d new variables.

(c) Separate Tests Based on Each Station Individually

An obvious strategy for using station information at m stations is to declare an event to be an outlier if any of the individual station-based tests finds the event to be an outlier. Fisk et al. (1995) and Gray et al. (1995) examined the use of a sequence of individual station tests with a Bonferroni-based adjustment to assure that the overall significance level is no larger than α.

Based on simulation studies, Fisk et al. (1995) and Gray et al. (1995) found that the power of the full-vector approach was consistently competitive with the other procedures. On the other hand, while the minimum variance weighting procedure could produce results with higher power than the full-vector approach in numerous cases, the results could be very poor since the weights were really not selected optimally for the purpose of outlier detection.

In Section 3, we derive station weights for the purposes of improving outlier detection. We also consider the use of a second-stage weighting procedure. Simulation results based on these techniques are also presented. In Section 4, we consider the problem of using multiple stations in outlier detection in which the population of the training sample is a mixture of component populations.

3. New Data Compression Procedures

In this section we consider two new procedures for data compression. While the full-vector approach considered in the previous section works very well, its implementation will be a problem when dimensionality becomes large. Once the monitoring procedure is operational, it may not be unusual for there to be as many as 5 to 10 stations that detect some events, and the numbered of features measured could realistically be as large as 8 or 9 so that the variance/covariance matrix could be 40×40 to 90×90. For this reason, it is desirable to

develop a data compression procedure, and in this section we propose compression techniques which give results that are similar to those of the full-vector approach.

(a) New station weights

For the kth feature we consider weights $\boldsymbol{\alpha_k} = (\alpha_{1k}, \ldots, \alpha_{mk})'$ to be determined as follows. Define the distance measure, $D_k(\alpha)$, between the compressed training set data, $\sum_{j=1}^{m} \alpha_{jk} \bar{X}_{jk}$, and the compressed potential outlier, $\sum_{j=1}^{m} \alpha_{jk} U_{jk}$, by

$$D_k(\alpha) = \frac{\{\boldsymbol{\alpha_k}'(\bar{\boldsymbol{X}}_k - \boldsymbol{U_k})\}^2}{(1+\frac{1}{n})\boldsymbol{\alpha_k}' \boldsymbol{S_k} \boldsymbol{\alpha_k}},$$

where $\boldsymbol{S_k}$ is the sample variance/covariance matrix of $\boldsymbol{X_{ki}}$. The $\widehat{\alpha}_{\boldsymbol{k}}$ that maximizes $D_k(\alpha)$ is $\widehat{\alpha}_k \propto \boldsymbol{S}_{\boldsymbol{k}}^{-1}(\bar{\boldsymbol{X}}_{\boldsymbol{k}} - \boldsymbol{U_k})$. These weights have the intuitively appealing feature that they maximize the distance between the potential outlier and the training data. Note that for each k, the compressed feature is univariate. This idea was originally used by Fisher (1936) in defining his now famous linear discriminant function.

We first consider the case in which $d = 1$, and we let $\widehat{\boldsymbol{\alpha}}' \boldsymbol{X}_{1i} = Y_{1i}$, $i = 1, \ldots, n$ and $\widehat{\boldsymbol{\alpha}}' \boldsymbol{U}_1 = \boldsymbol{V}_1$. Now, Y_{1i}, $i = 1, \ldots, n$ is a sample from some univariate distribution, and V_1 is a single observation from its univariate distribution. In the Appendix, we show that Hotelling's T^2 (actually the square of the Student's t in this univariate case) for testing $\mu_Y = \mu_V$ is numerically equivalent to the T^2 that would be calculated using (2) for the full-vector approach based on the original data.

Thus, we see that in the case $d = 1$, the use of $\widehat{\boldsymbol{\alpha}}$ weights produces a calculated T^2 value that is the same as that which would be calculated using the full-vector approach. Equivalently, the likelihood ratio, λ, is the same in each instance. Thus, if the distribution of λ is obtained using bootstrap techniques, the two procedures (full-vector on original data and compression using $\widehat{\boldsymbol{\alpha}}$) are equivalent. When $d > 1$ there is not a corresponding equivalence, but it is

intuitively expected that use of the $\widehat{\alpha}$ weights will produce compressed data that behaves in a manner similar to that of the full-vector approach and thus that use of the $\widehat{\alpha}$ weights will be preferable to the use of minimum variance weights $\widehat{\omega}$ in (3). One difference between the full-vector approach and the use of $\widehat{\alpha}$ weights for the case $d > 1$ is that the variance/covariance matrix for the compressed data is of order $d \times d$ as compared to an $md \times md$ variance/covariance matrix using the full-vector approach. It should be noted that since the results of the transformations (either using $\widehat{\alpha}$ or $\widehat{\omega}$) need not be normally distributed even though the original data may have been normal since the weights are not constant but are based on the data. For this reason, we recommend use of bootstrapping to obtain the appropriate critical value of λ for transformed data even if the original data were normal.

(b) A two-stage compression procedure

A second approach involves a second-stage compression across variables. We note that $\widehat{\alpha}_k'(\overline{X}_k - U_k)$ is a multiple of Hotelling's T^2 for the kth feature. We will denote this quantity as T_k^2. We then consider the random variable $Z = (T_1^2, \ldots, T_d^2)'$ and calculate

$$\begin{aligned} Q &= \max_{\beta} (\beta' Z)^2 / \beta' \widehat{\Sigma}_Z \beta \\ &= Z' \widehat{\Sigma}_Z^{-1} Z \end{aligned} \tag{4}$$

as an overall measure of how large the T_k^2's are where $\widehat{\Sigma}_Z$ is an estimate of Σ_Z and where (analogous to before) the weights that produce the maximum are $\widehat{\beta} = \widehat{\Sigma}_Z^{-1} Z$. It should be noted that large values of Q suggest that the observed value is an outlier, and we will use a bootstrap approach to approximate its distribution as before. Also, the original sample of size n produces only a single observation on Z, and because of this we use a separate bootstrap step to calculate $\widehat{\Sigma}_Z$. Specifically, we obtain B_1 nonparametric bootstrap samples of size $n + 1$ from the original training sample, and from each of these samples we calculate Z. We then let $\widehat{\Sigma}_Z$ be

the sample variance/covariance matrix. We then take B_2 nonparametric bootstrap samples of size $n + 1$ from the original training sample in order to find the null distribution of Q which is calculated as in (4) using the $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}}$ obtained from the first bootstrap step. Specifically, for purposes of the hypothesis test, the $100(1 - \alpha)$th percentile of $Q^*(b)$, $b = 1, ..., B_2$ is found. Ideally, given a bootstrap sample for which Q is to be calculated, a second bootstrap sample would be taken from this sample in order to obtain a bootstrap-based estimate of $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$ specific to that sample. However, this procedure would be very computationally intensive, and we have thus chosen the faster method of simply calculating $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}}$ once and using this estimate for each of the B_2 bootstrap samples. We will examine its performance using simulations.

(c) Simulation Results

In this section we show simulation results associated with the testing procedures in Sections 2 and 3. In particular, we consider the case in which there are $m = 2$ stations and $k = 2$ features. That is, the training sample consists of data values $\boldsymbol{X_i} = (X_{11i}, X_{12i}, X_{21i}, X_{22i})'$, $i = 1, ..., n$ where as in Section 2*a* the first and second subscripts represents station and feature respectively. In the simulations we simulated a training sample of size $n = 60$ along with a single observation from the specified "outlier population." Each test is utilized to determine whether the observation from the outlier population is determined to be an outlier. One thousand repetitions of this process were obtained, and the estimated power, i.e. proportion of times the outlier was detected, is given in Table 1 for each test for a variety of multistation scenarios. The 4×4 variance/covariance matrices associated with $\boldsymbol{X}$ are shown in the table. Each training sample mean is $(0, 0, 0, 0)'$ and in our simulations, the variance/covariance of the outlier population is taken to be the same as that of the population of the training sample. For each

TABLE 1. Significance Level and Power Results
$n = 60$, 1000 replications

	Outlier Mean Configuration	Combined Individual	Full Vector	MinVar Weights	New Weights	Double Weighting	Variance/Covariance Matrix
	1	.050	.044	.059	.050	.065	
	2	.601	.571	.388	.612	.546	
	3	.580	.547	.418	.573	.564	
A	4	.495	.551	.686	.579	.588	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 1 & 0 \\ & & & 1 \end{pmatrix}$
	5	.481	.571	.709	.599	.581	
	6	.813	.873	.949	.888	.858	
	1	.052	.045	.058	.063	.054	
	2	.170	.170	.086	.184	.178	
	3	.573	.544	.570	.567	.527	
B	4	.320	.352	.480	.365	.377	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 4 & 0 \\ & & & 4 \end{pmatrix}$
	5	.343	.364	.492	.405	.399	
	6	.629	.644	.783	.684	.674	
	1	.058	.051	.051	.058	.063	
	2	.355	.343	.265	.364	.400	
	3	.359	.341	.251	.367	.367	
C	4	.142	.147	.198	.166	.166	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ & 4 & 0 & 0 \\ & & 1 & 0 \\ & & & 4 \end{pmatrix}$
	5	.480	.520	.684	.554	.622	
	6	.591	.671	.789	.693	.677	
	1	.047	.057	.060	.064	.059	
	2	.585	.718	.314	.730	.699	
D	3	.600	.697	.336	.724	.665	$\begin{pmatrix} 1 & 0 & .5 & 0 \\ & 1 & 0 & .5 \\ & & 1 & 0 \\ & & & 1 \end{pmatrix}$
	4	.435	.376	.512	.406	.416	
	5	.424	.374	.493	.410	.432	
	6	.747	.693	.828	.716	.674	
	1	.046	.048	.053	.053	.058	
	2	.544	.896	.249	.913	.910	
E	3	.559	.915	.273	.919	.903	$\begin{pmatrix} 1 & 0 & .75 & 0 \\ & 1 & 0 & .75 \\ & & 1 & 0 \\ & & & 1 \end{pmatrix}$
	4	.375	.315	.422	.332	.379	
	5	.392	.350	.464	.363	.360	
	6	.683	.589	.744	.633	.605	
	1	.045	.048	.049	.052	.067	
	2	.347	.339	.241	.417	.495	
	3	.392	.398	.259	.459	.477	
F	4	.552	.512	.665	.535	.492	$\begin{pmatrix} 1 & .5 & .25 & 0 \\ & 1 & 0 & .25 \\ & & 1 & .5 \\ & & & 1 \end{pmatrix}$
	5	.558	.527	.663	.548	.469	
	6	.612	.627	.753	.699	.668	
	1	.044	.044	.053	.048	.079	
	2	.478	.551	.255	.604	.610	
	3	.465	.521	.269	.573	.621	
G	4	.456	.391	.525	.425	.453	$\begin{pmatrix} 1 & .25 & .5 & 0 \\ & 1 & 0 & 0 \\ & & 1 & .25 \\ & & & 1 \end{pmatrix}$
	5	.435	.381	.507	.416	.402	
	6	.645	.614	.767	.670	.651	

variance/covariance setting, six different mean configurations are considered for the potential outlier:

1: $(0, 0, 0, 0)'$

2: $(0, 0, 2, 2)'$

3: $(2, 2, 0, 0)'$

4: $(0, 2, 0, 2)'$

5: $(2, 0, 2, 0)'$

6: $(2, 2, 2, 2)'$

Obviously, mean configuration 1 is simply the mean of the training sample population so that estimated "power" results in this case are actually estimates of the significance level of the test. All tests were run at the nominal $\alpha = 0.05$ level, and all bootstrap procedures were based on 199 replications. In the table, it can be seen that all tests produced reasonable significance level results for all variance/covariance scenarios considered and that the full vector, maximum separation weights and the double weighting procedures all produced similar results. In fact, for several cases, the use of maximum separation weights and the double weighting produced power results higher than those for the full vector approach. As was observed previously, the minimum variance weights can at times produce power results that are larger than those of the three schemes just described. However, the power results using minimum variance weights can be very poor. See, for example, mean configurations 2 and 3 in scenarios D and E in the table. In these situations it is seen that the "combined individual" (i.e. separate tests using Bonferroni adjustment) tests are also inferior to the new weighting procedures given here since these weighting procedures take advantage of the correlation structure in the population of the training sample when testing a potential outlier.

4. Multistation Data with Mixture Training Samples

In this section we consider the scenario in which there may be more than one type of non-nuclear event in a region. Wang et al. (1996) consider this situation by assuming that the training data is a sample of size n from a mixture distribution whose density is given by

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} p_i g_i(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{1}$$

where m is the number of components in the mixture, $g_i(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the density associated with the ith component, the p_i, $i = 1, \ldots, m$ are the mixing proportions, and $\boldsymbol{x}$ is a d-dimensional vector of feature variables. For example, the mixture population might consist of events associated with earthquakes and mining explosions. The authors developed a modified likelihood ratio statistic, W, and a related test that required no distributional assumptions concerning the outlier distribution. The distribution of W under the null is unknown, so bootstrap procedures were developed to find an appropriate α-level critical value. The statistic W was calculated in such a way that small values of W were suggestive of an outlier. Let n_i, $i = 1, \ldots, m$ denote the sample sizes from each of the component populations so that $\sum_{i=1}^{m} n_i = n$. It is assumed that the training sample is selected in such a way that the n_i's contain information about the mixing proportions. Wang et al. (1996) assumed that some of the data were labeled, i.e. that the source components for these values are known. Sain, Gray, and Woodward (1996) have extended these results to the cases in which all data are unlabeled and to the case in which even the number of components is unknown. In this section we consider two multistation approaches for the mixture model case.

(a) A mixture-model approach

Recall that in Section 2b we discussed an approach that was based on first finding "optimal" station weights and then weighting across variables (i.e. a weighting of T^2 values). Unfortunately, in the mixture setting we cannot find the weights analogous to $\widehat{\boldsymbol{\alpha}}$ for optimally

combining stations for purposes of separating the outlier population from the population of the training sample. However, for a given feature k, one can consider the readings from the m stations to be an m-dimensional "feature" vector on which the modified likelihood ratio test of Wang et al. (1996) can be applied to obtain $W(k)$ which can be thought of as the result of combining station information somewhat analogous to obtaining a linear combination of the data from the m stations in the non-mixture case. Now, $1/W(k)$ behaves like T^2in that large values are suggestive of an outlier. A procedure somewhat analogous to that in Section 3*b* is to let $\boldsymbol{H} = (1/W(1), \ldots, 1/W(d))'$and calculate $D_H = \boldsymbol{H}'\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{H}}^{-1}\boldsymbol{H}$ as an overall measure of how large the $1/\mathrm{W}(k)'s$ are, where again, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{H}}$ is an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{H}}$ and must be calculated using a second bootstrap step.

(b) Closest component approach

A second possible approach for the mixture case would be to locate the component in the mixture that is "closest" to the outlier. Then, non-mixture techniques such as those given in Sections 2 and 3 could be used to determine whether the potential outlier should be considered to be an outlier from this closest component, and thus an outlier from the mixture.

5. Concluding Remarks

We have seen that in the non-mixture training sample scenario, the new weighting techniques produce power results that are competitive and sometimes better than those for the full-vector approach in all cases considered. Thus, we recommend the use of these new procedures when the number of stations and features is large. The procedure described in Section 4 for the mixture model case is currently being investigated using simulations. These results will be reported at a later time.

Appendix

In this Appendix we assume that $d = 1$ and that the random variables $Y_{1i} = \widehat{\boldsymbol{\alpha}}\,'\boldsymbol{X}_{1i}$, $i = 1, \ldots n$ and $V_1 = \widehat{\boldsymbol{\alpha}}\,'\boldsymbol{U}_1$ where $\widehat{\alpha}_k = \boldsymbol{S}_{\boldsymbol{k}}^{-1}(\,\overline{\boldsymbol{X}}_{\boldsymbol{k}} - \boldsymbol{U}_{\boldsymbol{k}})$ as developed in Section 3*a*. In this case, Hotelling's T^2 for testing $\mu_Y = \mu_V$ based on the transformed data is numerically equivalent to the T^2that would be calculated using (2) for the full vector approach using the original data. Now, Hotelling's T_Y^2 (actually the square of Student's t) for testing $U_Y = U_V$ based on the transformed data is

$$T_Y^2 = \left(\tfrac{n}{n+1}\right)(\overline{Y}_1 - V_1)' S_Y^{-1}(\overline{Y}_1 - V_1)\,.$$

where S_Y is the sample variance of Y_{1i}, $i = 1, \ldots, n$. Now, $S_Y = \widehat{\alpha}'\boldsymbol{S}\widehat{\alpha}$ where $\boldsymbol{S}$ is the sample variance/covariance matrix of the original data based on the training sample alone and $\overline{Y}_1 - V_1 = \widehat{\alpha}'(\overline{\boldsymbol{X}}_1 - \boldsymbol{U}_1)$. Thus, Hotelling's T_Y^2 is given by

$$T_Y^2 = \left(\tfrac{n}{n+1}\right)\widehat{\alpha}'(\overline{\boldsymbol{X}}_1 - \boldsymbol{U}_1)'(\widehat{\alpha}'\boldsymbol{S}\widehat{\boldsymbol{\alpha}})^{-1}\widehat{\boldsymbol{\alpha}}'(\overline{\boldsymbol{X}}_1 - \boldsymbol{U}_1)$$

$$=$$

$$\left(\tfrac{n}{n+1}\right)\left\{\left[(\overline{\boldsymbol{X}}_1 - \boldsymbol{U}_1)'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \boldsymbol{U})\right]\left[(\overline{X}_1 - \boldsymbol{U}_1)'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \boldsymbol{U}_1)\right]^{-1} \times\right.$$

$$\left.\left[(\overline{X}_1 - \boldsymbol{U}_1)'\boldsymbol{S}^{-1}(\overline{X}_1 - \boldsymbol{U}_1)\right]\right\}$$

$$= \left(\tfrac{n}{n+1}\right)\left[(\overline{\boldsymbol{X}}_1 - \boldsymbol{U}_1)'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \boldsymbol{U})\right]$$

which is Hotelling's T^2 for the original data as given in (2).

REFERENCES

Baek, J., Gray, H.L., McCartor, G.D., and Woodward, W.A.(1992). "A Generalized Likelihood Ratio Test in Outlier Detection or Script Matching," Advanced Research Projects Agency Technical Report.

Fisher, R.A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics* 7, 179-188.

Fisk, M., Gray, H.L., and McCartor, G.D. (1995), "Statistical Methodology and Assessment of Seismic Event Characterization Capability," Mission Research Corporation, Phillips Lab PL-TR-95-2156, ADA305487.

Gray, H.L., Woodward, W.A., and Yücel, Z.T. (1995), "Outliers with Multiple Stations," Southern Methodist University Technical Report, Department of Statistical Science, February, 1995.

Miller, J.W., Gray, H.L., and Woodward, W.A. (1993), "Discriminant Analysis and Outlier Testing when Data are Missing," Semi-Annual Technical Report Four, Advanced Research Projects Agency, Nuclear Monitoring Research Office, August 19, 1993.

Sain, S.R., Gray, H.L., and Woodward, W.A. (1996), "Outlier Detection without Ground Truth," *Proceedings of the18th Seismic Research Symposium on Monitoring a Comprehensive Test Ban Treaty,* PL-TR-96-2153, ADA313692.

Wang, S., Woodward, W.A., Gray, H.L., Wiechecki, S., and Sain, S.R. (1996), "A New Test for Outlier Detection from a Multivariate Mixture Distribution," *Journal of Computational and Graphical Statistics*, Submitted.

THOMAS AHRENS
SEISMOLOGICAL LABORATORY 252-21
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CA 91125

SHELTON ALEXANDER
PENNSYLVANIA STATE UNIVERSITY
DEPARTMENT OF GEOSCIENCES
537 DEIKE BUILDING
UNIVERSITY PARK, PA 16801

RICHARD BARDZELL
ACIS
DCI/ACIS
WASHINGTON, DC 20505

DOUGLAS BAUMGARDT
ENSCO INC.
5400 PORT ROYAL ROAD
SPRINGFIELD, VA 22151

WILLIAM BENSON
NAS/COS
ROOM HA372
2001 WISCONSIN AVE. NW
WASHINGTON, DC 20007

ROBERT BLANDFORD
AFTAC
1300 N. 17TH STREET
SUITE 1450
ARLINGTON, VA 22209-2308

RHETT BUTLER
IRIS
1616 N. FORT MEYER DRIVE
SUITE 1050
ARLINGTON, VA 22209

CATHERINE DE GROOT-HEDLIN
SCRIPPS INSTITUTION OF OCEANOGRAPHY
UNIVERSITY OF CALIFORNIA, SAN DIEGO
INSTITUTE OF GEOPHYSICS AND PLANETARY PHYSICS
LA JOLLA, CA 92093

SEAN DORAN
ACIS
DCI/ACIS
WASHINGTON , DC 20505

RICHARD J. FANTEL
BUREAU OF MINES
DEPT OF INTERIOR, BLDG 20
DENVER FEDERAL CENTER
DENVER, CO 80225

RALPH ALEWINE
NTPO
1901 N. MOORE STREET, SUITE 609
ARLINGTON, VA 22209

MUAWIA BARAZANGI
INSTITUTE FOR THE STUDY OF THE CONTINENTS
3126 SNEE HALL
CORNELL UNIVERSITY
ITHACA, NY 14853

T.G. BARKER
MAXWELL TECHNOLOGIES
P.O. BOX 23558
SAN DIEGO, CA 92123

THERON J. BENNETT
MAXWELL TECHNOLOGIES
11800 SUNRISE VALLEY DRIVE SUITE 1212
RESTON, VA 22091

JONATHAN BERGER
UNIVERSITY OF CA, SAN DIEGO
SCRIPPS INSTITUTION OF OCEANOGRAPHY IGPP, 0225
9500 GILMAN DRIVE
LA JOLLA, CA 92093-0225

STEVEN BRATT
NTPO
1901 N. MOORE STREET, SUITE 609
ARLINGTON, VA 22209

LESLIE A. CASEY
DOE
1000 INDEPENDENCE AVE. SW
NN-40
WASHINGTON, DC 20585-0420

STANLEY DICKINSON
AFOSR
110 DUNCAN AVENUE, SUITE B115
BOLLING AFB
WASHINGTON, D.C. 20332-001

DIANE I. DOSER
DEPARTMENT OF GEOLOGICAL SCIENCES
THE UNIVERSITY OF TEXAS AT EL PASO
EL PASO, TX 79968

JOHN FILSON
ACIS/TMG/NTT
ROOM 6T11 NHB
WASHINGTON, DC 20505

MARK D. FISK
MISSION RESEARCH CORPORATION
735 STATE STREET
P.O. DRAWER 719
SANTA BARBARA, CA 93102-0719

LORI GRANT
MULTIMAX, INC.
311C FOREST AVE. SUITE 3
PACIFIC GROVE, CA 93950

I. N. GUPTA
MULTIMAX, INC.
1441 MCCORMICK DRIVE
LARGO, MD 20774

JAMES HAYES
NSF
4201 WILSON BLVD., ROOM 785
ARLINGTON, VA 22230

MICHAEL HEDLIN
UNIVERSITY OF CALIFORNIA, SAN DIEGO
SCRIPPS INSTITUTION OF OCEANOGRAPHY IGPP, 0225
9500 GILMAN DRIVE
LA JOLLA, CA 92093-0225

EUGENE HERRIN
SOUTHERN METHODIST UNIVERSITY
DEPARTMENT OF GEOLOGICAL SCIENCES
DALLAS, TX 75275-0395

VINDELL HSU
HQ/AFTAC/TTR
1030 S. HIGHWAY A1A
PATRICK AFB, FL 32925-3002

RONG-SONG JIH
PHILLIPS LABORATORY
EARTH SCIENCES DIVISION
29 RANDOLPH ROAD
HANSCOM AFB, MA 01731-3010

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-200
LIVERMORE, CA 94551

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-221
LIVERMORE, CA 94551

ROBERT GEIL
DOE
PALAIS DES NATIONS, RM D615
GENEVA 10, SWITZERLAND

HENRY GRAY
SMU STATISTICS DEPARTMENT
P.O. BOX 750302
DALLAS, TX 75275-0302

DAVID HARKRIDER
PHILLIPS LABORATORY
EARTH SCIENCES DIVISION
29 RANDOLPH ROAD
HANSCOM AFB, MA 01731-3010

THOMAS HEARN
NEW MEXICO STATE UNIVERSITY
DEPARTMENT OF PHYSICS
LAS CRUCES, NM 88003

DONALD HELMBERGER
CALIFORNIA INSTITUTE OF TECHNOLOGY
DIVISION OF GEOLOGICAL & PLANETARY SCIENCES
SEISMOLOGICAL LABORATORY
PASADENA, CA 91125

ROBERT HERRMANN
ST. LOUIS UNIVERSITY
DEPARTMENT OF EARTH & ATMOSPHERIC SCIENCES
3507 LACLEDE AVENUE
ST. LOUIS, MO 63103

ANTHONY IANNACCHIONE
BUREAU OF MINES
COCHRANE MILL ROAD
PO BOX 18070
PITTSBURGH, PA 15236-9986

THOMAS JORDAN
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
EARTH, ATMOSPHERIC & PLANETARY SCIENCES
77 MASSACHUSETTS AVENUE, 54-918
CAMBRIDGE, MA 02139

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-207
LIVERMORE, CA 94551

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
LLNL
PO BOX 808, MS L-175
LIVERMORE, CA 94551

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-208
LIVERMORE, CA 94551

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-202
LIVERMORE, CA 94551

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-195
LIVERMORE, CA 94551

LAWRENCE LIVERMORE NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 808, MS L-205
LIVERMORE, CA 94551

THORNE LAY
UNIVERSITY OF CALIFORNIA, SANTA CRUZ
EARTH SCIENCES DEPARTMENT
EARTH & MARINE SCIENCE BUILDING
SANTA CRUZ, CA 95064

ANATOLI L. LEVSHIN
DEPARTMENT OF PHYSICS
UNIVERSITY OF COLORADO
CAMPUS BOX 390
BOULDER, CO 80309-0309

DONALD A. LINGER
DNA
6801 TELEGRAPH ROAD
ALEXANDRIA, VA 22310

LOS ALAMOS NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 1663, MS F659
LOS ALAMOS, NM 87545

LOS ALAMOS NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 1663, MS F665
LOS ALAMOS, NM 87545

LOS ALAMOS NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 1663, MS D460
LOS ALAMOS, NM 87545

LOS ALAMOS NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 1663, MS C335
LOS ALAMOS, NM 87545

GARY MCCARTOR
SOUTHERN METHODIST UNIVERSITY
DEPARTMENT OF PHYSICS
DALLAS, TX 75275-0395

KEITH MCLAUGHLIN
MAXWELL TECHNOLOGIES
P.O. BOX 23558
SAN DIEGO, CA 92123

BRIAN MITCHELL
DEPARTMENT OF EARTH & ATMOSPHERIC SCIENCES
ST. LOUIS UNIVERSITY
3507 LACLEDE AVENUE
ST. LOUIS, MO 63103

RICHARD MORROW
USACDA/IVI
320 21ST STREET, N.W.
WASHINGTON, DC 20451

JOHN MURPHY
MAXWELL TECHNOLOGIES
11800 SUNRISE VALLEY DRIVE SUITE 1212
RESTON, VA 22091

JAMES NI
NEW MEXICO STATE UNIVERSITY
DEPARTMENT OF PHYSICS
LAS CRUCES, NM 88003

JOHN ORCUTT
INSTITUTE OF GEOPHYSICS AND PLANETARY PHYSICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CA 92093

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K6-48
RICHLAND, WA 99352

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K7-34
RICHLAND, WA 99352

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K6-40
RICHLAND, WA 99352

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K5-72
RICHLAND, WA 99352

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K5-12
RICHLAND, WA 99352

KEITH PRIESTLEY
DEPARTMENT OF EARTH SCIENCES
UNIVERSITY OF CAMBRIDGE
MADINGLEY RISE, MADINGLEY ROAD
CAMBRIDGE, CB3 OEZ UK

PAUL RICHARDS
COLUMBIA UNIVERSITY
LAMONT-DOHERTY EARTH OBSERVATORY
PALISADES, NY 10964

CHANDAN SAIKIA
WOOODWARD-CLYDE FEDERAL SERVICES
566 EL DORADO ST., SUITE 100
PASADENA, CA 91101-2560

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 6116
MS 0750, PO BOX 5800
ALBUQUERQUE, NM 87185-0750

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 9311
MS 1159, PO BOX 5800
ALBUQUERQUE, NM 87185-1159

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 5736
MS 0655, PO BOX 5800
ALBUQUERQUE, NM 87185-0655

THOMAS SERENO JR.
SCIENCE APPLICATIONS INTERNATIONAL
CORPORATION
10260 CAMPUS POINT DRIVE
SAN DIEGO, CA 92121

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K7-22
RICHLAND, WA 99352

PACIFIC NORTHWEST NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
PO BOX 999, MS K6-84
RICHLAND, WA 99352

FRANK PILOTTE
HQ/AFTAC/TT
1030 S. HIGHWAY A1A
PATRICK AFB, FL 32925-3002

JAY PULLI
RADIX SYSTEMS, INC.
6 TAFT COURT
ROCKVILLE, MD 20850

DAVID RUSSELL
HQ AFTAC/TTR
1030 SOUTH HIGHWAY A1A
PATRICK AFB, FL 32925-3002

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 5704
MS 0979, PO BOX 5800
ALBUQUERQUE, NM 87185-0979

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 5791
MS 0567, PO BOX 5800
ALBUQUERQUE, NM 87185-0567

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 5704
MS 0655, PO BOX 5800
ALBUQUERQUE, NM 87185-0655

SANDIA NATIONAL LABORATORY
ATTN: TECHNICAL STAFF (PLS ROUTE)
DEPT. 6116
MS 0750, PO BOX 5800
ALBUQUERQUE, NM 87185-0750

AVI SHAPIRA
SEISMOLOGY DIVISION
THE INSTITUTE FOR PETROLEUM RESEARCH AND
GEOPHYSICS
P.O.B. 2286, NOLON 58122 ISRAEL

ROBERT SHUMWAY
410 MRAK HALL
DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616-8671

DAVID SIMPSON
IRIS
1616 N. FORT MEYER DRIVE
SUITE 1050
ARLINGTON, VA 22209

BRIAN SULLIVAN
BOSTON COLLEGE
INSITUTE FOR SPACE RESEARCH
140 COMMONWEALTH AVENUE
CHESTNUT HILL, MA 02167

NAFI TOKSOZ
EARTH RESOURCES LABORATORY, M.I.T.
42 CARLTON STREET, E34-440
CAMBRIDGE, MA 02142

GREG VAN DER VINK
IRIS
1616 N. FORT MEYER DRIVE
SUITE 1050
ARLINGTON, VA 22209

TERRY WALLACE
UNIVERSITY OF ARIZONA
DEPARTMENT OF GEOSCIENCES
BUILDING #77
TUCSON, AZ 85721

JAMES WHITCOMB
NSF
NSF/ISC OPERATIONS/EAR-785
4201 WILSON BLVD., ROOM785
ARLINGTON, VA 22230

JIAKANG XIE
COLUMBIA UNIVERSITY
LAMONT DOHERTY EARTH OBSERVATORY
ROUTE 9W
PALISADES, NY 10964

OFFICE OF THE SECRETARY OF DEFENSE
DDR&E
WASHINGTON, DC 20330

TACTEC
BATTELLE MEMORIAL INSTITUTE
505 KING AVENUE
COLUMBUS, OH 43201 (FINAL REPORT)

MATTHEW SIBOL
ENSCO, INC.
445 PINEDA COURT
MELBOURNE, FL 32940

JEFFRY STEVENS
MAXWELL TECHNOLOGIES
P.O. BOX 23558
SAN DIEGO, CA 92123

DAVID THOMAS
ISEE
29100 AURORA ROAD
CLEVELAND, OH 44139

LAWRENCE TURNBULL
ACIS
DCI/ACIS
WASHINGTON, DC 20505

FRANK VERNON
UNIVERSITY OF CALIFORNIA, SAN DIEGO
SCRIPPS INSTITUTION OF OCEANOGRAPHY IGPP, 0225
9500 GILMAN DRIVE
LA JOLLA, CA 92093-0225

DANIEL WEILL
NSF
EAR-785
4201 WILSON BLVD., ROOM 785
ARLINGTON, VA 22230

RU SHAN WU
UNIVERSITY OF CALIFORNIA SANTA CRUZ
EARTH SCIENCES DEPT.
1156 HIGH STREET
SANTA CRUZ, CA 95064

JAMES E. ZOLLWEG
BOISE STATE UNIVERSITY
GEOSCIENCES DEPT.
1910 UNIVERSITY DRIVE
BOISE, ID 83725

DEFENSE TECHNICAL INFORMATION CENTER
8725 JOHN J. KINGMAN ROAD
FT BELVOIR, VA 22060-6218 (2 COPIES)

PHILLIPS LABORATORY
ATTN: XPG
29 RANDOLPH ROAD
HANSCOM AFB, MA 01731-3010

PHILLIPS LABORATORY
ATTN: GPE
29 RANDOLPH ROAD
HANSCOM AFB, MA 01731-3010

PHILLIPS LABORATORY
ATTN: TSML
5 WRIGHT STREET
HANSCOM AFB, MA 01731-3004

PHILLIPS LABORATORY
ATTN: PL/SUL
3550 ABERDEEN AVE SE
KIRTLAND, NM 87117-5776 (2 COPIES)